



## **Portfolio Review Committee Agenda**

December 17, 1:00 p.m. to 3:00 p.m.  
520 Mendocino Ave, Santa Rosa, Laurel Room

*All supporting documents are available at [www.UpstreamInvestments.org](http://www.UpstreamInvestments.org) and at the Board of Supervisors office at 575 Administration Drive, Room 100A, Santa Rosa, CA, during normal business hours. For accessibility assistance with this agenda or supporting documents, please e-mail [Upstream@schsd.org](mailto:Upstream@schsd.org) or call 707.565.5800.*

- 1:00 Welcome, Introductions, Minutes ACTION ITEM**  
*Review and approve November minutes*
- 1:05 Updates & Reports**  
*Update on portfolio activity since last meeting*
- 1:15 Literature Review Cover Form ACTION ITEM**
- 1:25 SAMHSA Appropriateness of Analysis ACTION ITEM**
- 1:45 History and Impact of OJJDP Changes ACTION ITEM**
- 2:10 Establish Tier 2 Renewal Procedures ACTION ITEM**
- 2:55 Public Comment**
- 3:00 Adjourn**

### **Upcoming Meetings**

January 21, 2015, February 18, 2015, March 18, 2015  
April 15, 2015, May 20, 2015, June 17, 2015  
July 15, 2015, August 19, 2015, September 16, 2015,  
October 21, 2015, November 18, 2015, December 16, 2015

All meetings will be held from 1:00 – 3:00 pm at 490 Mendocino Ave in the Manzanita Room unless otherwise noted



**Wednesday, November 19, 2014 Portfolio Review Committee Meeting Minutes**

**1:00 p.m. – 3:00 p.m.**

**520 Mendocino Ave, Santa Rosa, Laurel Conference Room**

**Members (listed alphabetically)**

BJ Bischoff, Bischoff Consulting  
 Carlos Ayala, Sonoma State University  
 Carol Simmons, Child Care Planning Council  
 Jennifer O’Donnell, United Way  
 Julie Sabbag-Maskey, Human Services Department  
 Kate Pack, First 5  
 Katie Greaves, Human Services Department  
 Leo Tacata, District Attorney’s Office  
 Monique Chapman, Sheriff’s Office  
 Rob Halverson, Probation Department  
 Serena Lienau, City of Santa Rosa

**Staff (listed alphabetically)**

Angie Dillon-Shore, Human Services Department  
 Joni Thacher, Human Services Department

**Not Present (listed alphabetically)**

Ellen Bauer, Department of Health  
 Rebecca Wachsberg, Probation Department  
 Stephen Jackson, SCOE

**Public (listed alphabetically)**

Cathryn Couch, Ceres Community Project  
 Debra Solomon, Human Services Agency of San Francisco  
 Kristin James-Bowe, Napa County Health and Human Services

<b>Topic</b>	<b>Discussion</b>	<b>Decision</b>	<b>Next Steps</b>
<b>Welcome, Introductions, Minutes, Updates</b>	Angie welcomed everyone and facilitated introductions. Motion to approve the minutes.  Angie provided an update of recent technical assistance and programs added to the Portfolio.	Motion to approve the minutes: BJ Second: Carol Yes: 12 No: 0 Abstain: 0	None
<b>Application to the Portfolio</b>	<b>Ceres Community Project Healthy Meals for Heathy Communities</b> <b>Items for discussion: Logic Model and Evaluation</b> <b>Youth component is not addressed but is a critical element of the program</b> B.J. and Carlos recused themselves Monique approved the program, Serena denied the program. Angie explained that Ceres submitted an application specifically for their Teen Mentoring program earlier this year. A copy of the program was provided for the committee’s reference.	Motion to approve Healthy Meals for Healthy Communities to the Portfolio and add a footnote to reference Ceres Teen Mentoring application.	Staff members will add a footnote. Healthy Meals will be included in the Upstream Portfolio.

Topic	Discussion	Decision	Next Steps
	<p><b>Monique:</b> I had no concerns and had also reviewed Cere’s Teen Mentoring program that was submitted earlier this year. This program differs in that it provides services to both adults and youth. This was an excellent and very robust application. My only concern was that there was a lot of information, and I didn’t want to miss anything. I felt it met our standards.</p> <p><b>Serena:</b> I felt the application was very well put together. My concern was that there was very little mention of the youth piece. I felt that this was a critical component that needed to be developed. I didn’t review the other application and it didn’t feel like a complete package. I would like a better understanding of why the programs were separated. This was a fantastic application, but I would like to see the youth piece developed.</p> <p><b>Monique:</b> I was initially confused. I thought it was the same application I’d reviewed earlier this year. I removed the youth piece from my thinking when I reviewed this application. The youth application was specific to mentoring, this application is specific to providing meals.</p> <p><b>Serena:</b> I think it’s great that the applications are separate, but I need to see linkage between the two. As simple as making a reference in the logic model, or a note to reference the other application.</p> <p><b>Committee Discussion:</b></p> <ul style="list-style-type: none"> <li>• Are they so separate they can stand on their own without a reference to each other? Could the programs exist on their own?</li> <li>• Maybe there just needs to be a footnote that references the other application?</li> </ul> <p><b>Angie:</b> Is there any more discussion before I invite public comment?</p> <p><b>Cathryn:</b> The client side is more important to the youth development than the inverse. It’s a complicated program to put into one application. Because we run them both separately we chose to pull them apart for the applications, but the truth is they are totally integrated. We had an assumption the committee would be aware of our previous application.</p> <p><b>Committee Discussion:</b></p> <ul style="list-style-type: none"> <li>• They really are one program. Maybe we could approve them as a part 1 and 2?</li> <li>• They have different strategies, outcomes, and evaluations though. We have other programs, like Triple P, that are part of a larger program.</li> <li>• If either of these can be standalone programs they should be separate applications.</li> <li>• The logic models are very different. I found enough evidence to demonstrate that these</li> </ul>	<p>Motion: Rob Second: Monique Yes : 10 No: 0 Abstain: 2</p>	

Topic	Discussion	Decision	Next Steps
	<p>are separate programs.</p> <ul style="list-style-type: none"> <li>• Maybe as these programs move from Tier 2 to Tier 3 the application could integrate both programs.</li> </ul> <p><b>Rob:</b> Move to approve Healthy Meals for Healthy Communities if a footnote is added to reference the Teen Mentoring application for more details on the youth component. Staff members can do this, it does not need to be returned to the applicant.</p> <p><b>Monique:</b> I second the motion.</p> <p><b>Healthy Meals for Healthy Communities will be included as a Tier 3 program in the Portfolio for 3 years. Staff will add a footnote to link the Teen Mentoring program.</b></p>		
<b>Additional days for a review</b>	<p>Rebecca submitted a request agenda item form asking the committee to consider increasing the time for reviews from 5 days to 10 days.</p> <p><b>Discussion:</b></p> <ul style="list-style-type: none"> <li>• Originally we only had 3 days to return a review. What was the justification for the short turn around?</li> <li>• Many applicants use this for funding. The faster the better. Five days was a compromise.</li> <li>• Funding is still a consideration, but most organizations have months to complete funding requirements. They need to plan ahead.</li> <li>• If you give me more days, I'll take more days and still be late.</li> <li>• Reminders help. It would be nice to get a half-way reminder.</li> <li>• Rebecca wasn't able to attend this meeting. She expressed that allowing for a 10 day review would ensure she had a weekend to complete the review. If there isn't a significant reason to complete the review in five days, it's not worth it to reconfigure her schedule to try and get it done.</li> <li>• Could we allow for requests for expedited reviews?</li> <li>• No, organizations need to plan ahead. Even if an application is submitted months in advance of a deadline, the process can still be lengthy if it goes to committee or needs TA.</li> </ul> <p><b>Review Committee members will now have 10 days to complete reviews.</b></p>	<p>Motion to increase review time to 10 days.  Motion: Carol  Second: Karin  Yes: 12  No: 0  Abstain: 0</p>	<p>Staff members will update the bylaws.</p>
<b>SAHMSA Required score &amp; Appropriateness Analysis</b>	<p>Our instructions require programs to be ranked 2 or 4 for all SAHMSA criteria, but scores fall between 2 and 4, e.g. 3.9, 2.5. This is confusing for applicants. Can we revise this to include anything 2 or above?</p> <p><b>Language will be revised to read "2 or higher."</b></p> <p>We currently require all programs to have a 4 for data analysis. Very few programs are ranked this high. Could we consider accepting lower scores?</p>	<p>Motion to accept 2 or higher for all criteria except Appropriateness of Analysis.  Motion: Katie</p>	<p>Staff members will research SAHMSA review standards.</p>

Topic	Discussion	Decision	Next Steps
	<p><b>Discussion:</b></p> <ul style="list-style-type: none"> <li>• Why did we elevate analysis?</li> <li>• It's elevated because it demonstrates a relationship between interventions and outcomes.</li> <li>• The language describing what a 2 means doesn't suggest anything positive for this category. It says the analysis may not have been appropriate or it was inadequate. In the other categories a 2 suggests something positive.</li> <li>• It would be good to know what a score between 2 and 4 means.</li> <li>• The spirit of the committee is to have a yes/no switch for clearinghouses. To construct a similar understanding the committee said everything is good on SAHMSA except for the evidence-based piece which must be a 4. Researchers are going to say this is a 0 or 4. We wanted something that was very solid and clear. I struggle with not having a 4 but understand that this could sink a lot of applications.</li> </ul> <p><b>B.J. motioned to accept a 2 or higher.</b></p> <p><b>Discussion:</b></p> <ul style="list-style-type: none"> <li>• That means that their sample size could be inadequate. Are we OK with that?</li> <li>• Couldn't we infer that if it's a 3 most reviewers thought it was a 4?</li> <li>• I'm concerned we do not want certain programs to be affected. We need to maintain rigor.</li> <li>• We need more time and information. We need to know how many reviewers, do reviewers assign a range of scores, what's the definition of a score between 2 and 4.</li> </ul> <p><b>Discussion tabled. Staff members will provide additional information in December.</b></p>	<p>Second: Serena Yes: 12 No: 0 Abstain: 0</p>	
<p><b>OJJDP changed rating scale</b></p>	<p>We require programs to be rated exemplary or effective. There is no longer and exemplary rating. The new scale includes effective, promising, and no effect. This will affect 4 tier 1 programs that are renewing and are currently rated promising on the OJJDP site.</p> <p><b>Discussion:</b></p> <ul style="list-style-type: none"> <li>• All programs have been rerated with the Crime Solutions rating scale.</li> <li>• The OJJDP definition of promising is "in general this study demonstrates promising (perhaps inconsistent) evidence in favor of the program when evaluated with a design of high quality (quasi-experimental). More extensive research is required."</li> <li>• This does not meet our standard for Tier 1.</li> <li>• All we can accept from OJJDP is effective.</li> </ul> <p><b>Carol motioned to only accept programs rated effective by the OJJDP for Tier 1 inclusion.</b></p> <p><b>Discussion:</b></p> <ul style="list-style-type: none"> <li>• There is a larger question. What do we do with programs that were Tier 1 and will no</li> </ul>	<p>Motion to only accept programs rated effective by the OJJDP for Tier 1 Portfolio inclusion. Motion: Carol Second: Julie Yes: 12 No: 0 Abstain: 0</p>	<p>Staff will contact renewing organizations that will be affected.</p>

Topic	Discussion	Decision	Next Steps
	<p>longer be Tier 1?</p> <ul style="list-style-type: none"> <li>• Funders don't give more points for a higher Tier. It's not a big deal.</li> <li>• This means that 4 organizations that have been Tier 1 programs will now have to complete a Tier 2 application to stay on the Portfolio. Three of these organizations have already completed and submitted their Tier 1 renewal applications.</li> <li>• This is why the Tier 1 renewal process is so important. Our website and instructions need to reflect that just because a program was on a clearinghouse doesn't mean it will always be on a clearinghouse.</li> <li>• Rating is always subject to change. This is a changing and growing field. It's a data driven culture.</li> </ul> <p>Affected organizations will need to complete a Tier 2 application.</p>		
<b>Accreditation</b>	<p>In May 2013 the Portfolio Review Committee decided not to accept accreditation for Tier 1 programs. All programs would need to complete a fidelity chart. However, the committee elected to leave a note on Tier 1 applications instructing organizations participating in an accreditation process to contact staff members before completing the fidelity chart. What should staff members response be when this happens?</p> <p><b>Discussion:</b></p> <ul style="list-style-type: none"> <li>• They could submit everything they're required to submit for accreditation and we could evaluate that.</li> <li>• What about a time limit? How recent does their accreditation have to be?</li> <li>• Everyone should have to complete the fidelity table.</li> </ul> <p><b>Decision to remove all references to accreditation and require all organizations to complete the fidelity table.</b></p>	Remove all references to accreditation.	Staff members will update application documents.
<b>Tier 2 Renewals</b>	<p>If there have been no changes to a program can staff approve Tier 2 renewals?</p> <ul style="list-style-type: none"> <li>• I thought we were going to see the original packet?</li> <li>• You don't need us if we're not re-evaluating a program.</li> <li>• If we re-review the original program we're second guessing our original decision.</li> <li>• We need to do that. We need to re-evaluate every Tier 2 program.</li> <li>• We need to look at when our process changed. They should be held to current standards.</li> <li>• We need a current list of all the changes we've made to our process over the years.</li> <li>• It's possible that even if the criteria have remained the same, there's a possibility that the new reviewers will have different opinions.</li> </ul> <p><b>Conversation tabled. We will discuss Tier 2 renewal procedures in December.</b></p>	None	Staff members will review historical changes.

<b>Topic</b>	<b>Discussion</b>	<b>Decision</b>	<b>Next Steps</b>
<b>Literature Review Cover Sheet</b>	Would it be helpful to have a standard literature review cover sheet that required organizations to articulate how the literature related to their programming? Some organizations submit multiple lengthy literature reviews. <b>Staff members will create a template and bring it to the December meeting.</b>	None	Staff members will create a cover sheet template.
<b>Public Comment</b>	None at this time.	None	None
<b>Next Meeting/Final Comments</b>	The next meeting will be on December 17 at 520 Mendocino Ave even if there are no applications for review.	None	None
<b>Adjourn</b>	Adjourned at 3:00 pm.	None	None



## Update Report for the Portfolio Review Committee

December 2014

**To date 78 programs have been approved for tier placement on the Portfolio:**

Tier 1: 25 programs

88 organizations *implementing* programs on the Portfolio

Tier 2: 29 programs

55 local organizations *funding* programs on the Portfolio

Tier 3: 24 programs

### New submissions since November 19, 2014

	Approval Date	Program	Tier	Submitting Agency	Received TA?
1	11/19/14	Healthy Meals for Healthy Communities	3	Ceres Community Project	Yes
2	11/20/14	Teen Parent Connections	2	Sonoma County Department of Health Services	Yes
3	Under Review	Maternal Child Health Field Nursing	3	Sonoma County Department of Health Services	No
4	Under Review	Motivational Interviewing	1	Social Advocates for Youth	Yes
5	Under Review	Motivational Interviewing	1	Community Action Partnership	Yes
6	Under Review	WIC Dental Days	2	Community Action Partnership	Yes
7	Under Review	Diabetes Self-Management Program <i>RENEWAL</i>	1	Northern California Center for Well-Being	Yes

### 1:1 TA provided to renewal and active submissions since November 19, 2014

	Submission Date	Program	Submitting Agency	Status
1	9/23/2011	Motivational Interviewing	DAAC	Tier 1 renewal TA
2	11/6/2011	Project RESPECT	DAAC	Tier 1 renewal TA
3	12/31/11	Project Success	DAAC	Tier 1 renewal TA
5	3/6/13	iDo26.2	Volunteer Center	Tier 3 to Tier 2 transition TA
6	12/5/13	Neighborhood Listening Project	Listening for a Change	Tier 2 denied, receiving TA
7	3/25/2014	Algebra Academy	Santa Rosa Chamber of Commerce	Tier 3 denied, receiving TA
8	8/13/14	TALKK	Sonoma County Office of Education	Tier 3 to Tier 2 transition TA

2014 total 1:1 technical assistance hours: **102.5**

2014 total presentations/workshops: **27**

Summary Report, Portfolio of Model Upstream Programs ([updated. 12.12.14](#))

For more information visit [www.UpstreamInvestments.org](http://www.UpstreamInvestments.org), contact [Upstream@schsd.org](mailto:Upstream@schsd.org) or call 565.5800.

### Proposed Pre-written Literature Review Cover Sheet

If you are not writing your own literature review, please complete a cover sheet that includes:

- 1) Introduction
  - a. Problem statement
  - b. Scope of the problem including incidence and prevalence
- 2) The citation for the literature review(s) you are submitting (See examples below).
- 3) For each Literature review, please summarize its original intended purpose and how it relates to your program, for example, what it supports, or doesn't support from the literature
- 4) Conclusion-connect all the lit reviews together in a coherent picture of how they support your entire program. Make sure it reflects your own understanding of the data presented in the articles you use.

#### Examples of Reference List

##### *Journal Article*

Rosler, A., Ulrich, C., Billino, J., Sterzer, P., Weidauer, S., Bernhardt, T., ...Kleinschmidt, A. (2005). Effects of arousing emotional scenes on the distribution of visuospatial attention: Changes with aging and early subcortical vascular dementia. *Journal of the Neurological Sciences*, 229, 109–116. doi:10.1016/j.jns.2004.11.007

##### *Manual*

Spielberger, C. D., Gorsuch, I., & Lushene, R. E. (1970). *Manual for the State–Trait Inventory*. Palo Alto, CA: Consulting Psychologists Press.

##### *Tool*

Wechsler, D. (1987). *Wechsler Memory Scale—Revised*. San Antonio, TX: Psychological Corporation.

##### *Book*

Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 1–90). Mahwah, NJ: Erlbaum

# CRIMESOLUTIONS.GOV: PROGRAM EVIDENCE RATING INSTRUMENT—PART 1

**Instructions:** Please carefully assess the program in terms of the conceptual framework. **The reviewer should complete Part 1 only once for each program**, regardless of the number of studies to be reviewed. Complete this section by using the pertinent information from the studies and any other program materials you have received. Please record your answers on this form.

**PROGRAM NAME:** \_\_\_\_\_

**REVIEWER'S NAME** \_\_\_\_\_

**DATE OF REVIEW** \_\_\_\_\_

## CONCEPTUAL FRAMEWORK

**A. PRIOR RESEARCH** assesses the degree to which previous empirical evidence (formal evaluations and meta-analyses) supports the conceptual framework of *comparable* programs. It is important to note that the scope of comparable programs will vary by program. For instance, Multisystemic Therapy (MST) has undergone numerous evaluations, thus the scope of comparable programs can be narrowed to consist solely of MST rather than include the other family-based treatment models. On the other hand, a program such as the Chicago Alternative Policing Strategy (CAPS) is unlikely to have been the subject of repeated evaluation. In this case, the scope of comparable programs can be widened to include other similar community-based policing programs. *(Notes: A meta-analysis will typically include five or more studies on a single practice. Consequently, if a meta-analysis provides evidence to support the program, the research base should receive the highest score. Also note: an independent evaluator is **NOT** required for consideration, but specify the association between the program and evaluator [if known].)*

CHECK	POINTS	DESCRIPTION
	3 =	High (5 or more other studies, or 1 meta-analysis, provide evidence in support of the program).
	2 =	Medium (2 to 4 other studies provide evidence in support of the program).
	1 =	Low (1 other study provides evidence in support of the program).
	0 =	None (No other studies provide evidence in support of the program).

Program-Evaluator Association: \_\_\_\_\_

Notes: \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**B. THEORETICAL BASE** measures the degree to which the program is based on a well-articulated, conceptually sound program theory. Some programs are designed with little regard to conceptual development other than an implicit appeal to common sense. Instead, a program should provide an explanation of why and how it is expected to achieve its intended results and should be supported by prior conceptual development and empirical research.

CHECK	POINTS	DESCRIPTION
	3 =	Program theory is fully described and conceptually sound.
	2 =	Program theory is adequately described and appears conceptually sound.
	1 =	Very little information is provided about program theory, but it may be conceptually sound.
	0 =	No information about program theory or program theory is invalid.

Notes: \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**C. PROGRAM DESCRIPTION** rates the degree to which the program details are described. A full and thorough description should serve as a guide for the implementation of the program. It would include the following information: 1) the logic of the program, 2) the details of all key components, 3) the frequency and duration of the program activities, 4) the targeted population, 5) the targeted behavior(s) (i.e., the intent of the program), and 6) the setting. The rating should reflect the degree to which the provided materials afford an adequate program description and/or direct the reader to references containing such a description.

CHECK	POINTS	DESCRIPTION
	3 =	All program details are specified.
	2 =	Most program details are specified.
	1 =	Some program details are specified.
	0 =	No program details are specified.

Notes: Please specify the targeted population, the targeted behaviors, and the key elements of the program:

---



---



---



---



---



---



---



---

CONCEPTUAL FRAMEWORK SCORING TABLE		
	Prior Research Points	
+	Theoretical Base Points	
+	Program Description Points	
=	TOTAL	
/	NUMBER OF ITEMS	3
=	CONCEPTUAL FRAMEWORK SCORE	

**CRIMESOLUTIONS.GOV: PROGRAM EVIDENCE RATING INSTRUMENT—PART 2**

**Instructions:** Please carefully assess the program in terms of design quality, outcome evidence, and program fidelity. **Part 2 should be completed for each study in the research base. Please record your answers for each article on this form.** (Note: The research base for each program can include up to three studies.)

<b>PROGRAM:</b>			
<b>STUDY #:</b>		<b>CITATION</b>	

**REVIEWER'S NAME** \_\_\_\_\_ **DATE OF REVIEW** \_\_\_\_\_

**DESIGN QUALITY**

**A. RESEARCH DESIGN** rates the ability of the design to infer a causal relationship between program treatment and outcome. There are three general types of designs: experimental, quasi-experimental, and non-experimental. The designs differ in the method of assignment. A randomized field experiment randomly sorts participants into two or more groups. One group receives the program (treatment), while the other (controls) does not<sup>1</sup>. A quasi-experiment research design is similar with the exception that the subjects are assigned to the treatment and comparison groups through a process that is not random. Finally, a non-experiment lacks one or both of the above characteristics. Since these designs differ in their assignment strategy, it is likely they will differ in terms of their strength with respect to internal validity. (Note: Not all designs easily fit into this hierarchy. The reviewer should specify the design and note the reason for the score.)

CHECK	POINTS	DESCRIPTION
	3 =	Experimental (well-designed randomized field trial).
	2 =	Quasi-experimental Level 1 (design uses a credible comparison group with extensive information provided on pre-treatment equivalence of groups; time series comparison group design).
	1 =	Quasi-experimental Level 2 (design has a comparison group but lacks comparability on important preexisting variables or lacks information on pre-treatment equivalence of groups; time series single group design).
	0 =	Non-experimental (one group pretest–posttest, one- and two-group posttest only, or case studies).

Specify Design: \_\_\_\_\_

Notes: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

<sup>1</sup> In some cases, random assignment takes place at a different level than the analysis. For example, schools are randomly assigned to conditions, but the students are the unit of analysis. These cases should not be treated as random assignments.



## CRIMESOLUTIONS.GOV: PROGRAM EVIDENCE RATING INSTRUMENT—PART 2

**B. SAMPLE SIZE (POWER)** assesses the adequacy of the sample to detect meaningful program effects. However, the optimal size of a sample is rarely straightforward. Statistical power is a function of several factors: 1) the size of the sample; 2) the magnitude of the expected effect; 3) the type of statistical test used; and 4) the alpha level set to control Type I error (conventionally set at .05). In general, for a traditional two group experiment with a statistical power of .80, the N should be roughly 394 per group to detect a small effect ( $d=.20$ ); 64 to detect a medium effect ( $d=.50$ ); and 26 to detect a large effect ( $d=.80$ ). It should be noted however that these figures are guidelines to help direct the review. (*Note 1: For three groups, the N per group drops to roughly 322 for a small effect, 52 for a medium effect, and 21 for a large effect. Group size continues to drop as the number of groups increases. Note 2: The same rules of thumb do not apply for time series designs. Most textbooks suggest that about 50 observations, with a reasonable distribution among pre- and posttest measurements, is required for a competent analysis, on grounds that this figure is usually sufficient for estimating the structure of the correlated error. Conversely, although it may not account for the randomness of the data, roughly 15 observations are generally considered the minimum.*) The reviewer should use his or her expertise to assess the adequacy of the sample.

CHECK	POINTS	DESCRIPTION
	3 =	High Power: The sample is sufficient to detect a small effect (.20) using appropriate tests. (In general, the N should be greater than 394 per group in a traditional experiment and greater than 75 in a time series design.)
	2 =	Medium Power: The sample is sufficient to detect a medium effect (.50) using appropriate tests. (In general, the N should be between 64 and 393 per group in a traditional experiment and between 51 and 75 in a time series design.)
	1 =	Low Power: The sample is sufficient to detect a large effect (.80) using appropriate tests. (In general, the N should be between 26 and 63 per group in a traditional experiment and between 15 and 50 in a time series.)
	0 =	Insufficient: The sample is not sufficient to detect an effect. (In general, the N is less than 25 per group in a traditional experiment and less than 15 in a time series design.)

Specify treatment group sample size:

Specify comparison group sample size:

Specify number of observations (Time Series design):

Notes:

**C. STATISTICAL ADJUSTMENT (if applicable)** assesses the use of statistical controls to account for the initial measured differences between the groups. Any outcome-relevant variable on which the groups may differ should be identified and included in the statistical adjustment. (*Note 1: Some program studies, such as place and field studies in situational crime prevention, do not lend themselves to the use of statistical controls. In such cases, please choose not applicable.*)

CHECK	POINTS	DESCRIPTION
	3 =	No statistical adjustments required in the analysis. Random assignment or selection modeling (propensity score matching) with a sufficiently large sample resulted in no group differences.
	2 =	The analysis employs appropriate statistical adjustments (includes control variables that are presumed to be related to the outcome) to control for group differences.
	1 =	The analysis employs statistical adjustments (includes control variables that are presumed to be related to the outcome) but some important variables are not addressed.
	0 =	The analysis does not employ necessary statistical adjustments to control for group differences.
	NA	Not applicable.

Notes:

**CRIMESOLUTIONS.GOV: PROGRAM EVIDENCE RATING INSTRUMENT—PART 2**

**D. INSTRUMENTATION** rates the quality (reliability and validity) of the measures used in the study. Reliability refers to the stability and consistency of the measures. Validity refers to the accuracy of the measure. The selection of appropriate instrumentation should also consider the developmental and cultural appropriateness of the measure, as well as the reading level, native language, and attention span of respondents.

CHECK	POINTS	DESCRIPTION
	3 =	Excellent. The reliability (the extent to which an item produces the same results when used repeatedly) and validity (the extent to which an item measures what it is intended to measure) of the measures are excellent.
	2 =	Adequate. The reliability (the extent to which an item produces the same results when used repeatedly) and validity (the extent to which an item measures what it is intended to measure) of the measures are adequate.
	1 =	Below Average. The reliability (the extent to which an item produces the same results when used repeatedly) and/or validity (the extent to which an item measures what it is intended to measure) of the measures are below average.
	0 =	None. No information is provided on the reliability (the extent to which an item produces the same results when used repeatedly) and/or validity (the extent to which an item measures what it is intended to measure) of the measures.

Notes:

---



---



---



---

**E. INTERNAL VALIDITY** assesses the degree to which the observed changes can be attributed to the program. The validity of a study depends on both the research design and the measurement of the program activities and outcomes. Threats to internal validity will affect the accuracy of the results and draw into question the effect of the intervention.

**Please check the specific threats to validity in the table on the next page and include notes.**

CHECK	POINTS	DESCRIPTION
	3 =	No threats to internal validity are identified or all threats have been adequately addressed.
	2 =	Marginal threats to internal validity are identified and remain.
	1 =	Moderate threats to internal validity are identified and remain.
	0 =	Serious threats to internal validity are identified and remain.

Notes:

---



---



---



---

**CRIMESOLUTIONS.GOV: PROGRAM EVIDENCE RATING INSTRUMENT—PART 2**

Check all that apply	Threat	Description
<input type="checkbox"/>	Attrition or Mortality	<p>This threat occurs when participants drop out of the study between the pretest and the posttest. Attrition is important because it affects whether the groups are equivalent except for program effects at the time of the post-program outcome measure. The study should have low overall attrition of study participants and minimal differential attrition between the treatment and control groups. While there are exceptions, the general guideline states that a study should obtain outcome data for at least 80 percent of the original study subjects. Furthermore, the attrition rate should be approximately the same for the treatment and control groups. Severe differential attrition makes the results suspect, because it may compromise the comparability of the groups.</p> <p>Notes: _____            _____            _____</p>
<input type="checkbox"/>	Maturation	<p>This threat is caused by the natural maturation process, where respondents grow experienced or bored.</p> <p>Notes: _____            _____            _____</p>
<input type="checkbox"/>	Instrumentation	<p>This threat occurs when there is a change in the measuring instrument.</p> <p>Notes: _____            _____            _____</p>
<input type="checkbox"/>	Regression Toward the Mean	<p>This threat occurs whenever there is measurement error and participants are selected based on the extremeness of their measured values. The measured values will tend to be closer to the overall mean on a second administration of the instrument.</p> <p>Notes: _____            _____            _____</p>
<input type="checkbox"/>	Selection	<p>This threat occurs when the groups to be compared differ on factors besides the treatment. Even if the subjects are randomly assigned, this threat is of particular importance with small sample studies.</p> <p>Notes: _____            _____            _____</p>
<input type="checkbox"/>	Contamination	<p>This threat refers to situations where the separation between the groups is less than it should be.</p> <p>Notes: _____            _____            _____</p>
<input type="checkbox"/>	History	<p>This threat occurs when an observed effect might be due to an event that takes place between the pretest and the posttest that has nothing to do with the treatment.</p> <p>Notes: _____            _____            _____</p>
<input type="checkbox"/>	Other	<p>Other threats may include: multiple treatment interference, obtrusive testing, secular trends, intervening events, etc.</p> <p>Notes: _____            _____            _____</p>

**CRIMESOLUTIONS.GOV: PROGRAM EVIDENCE RATING INSTRUMENT—PART 2**

**F. FOLLOW-UP PERIOD** assesses the length of time that the study period continues after the program ends to ascertain its sustained effects. In cases where programs do not have clearly defined endpoints, the follow-up period may be delimited between the first and last assessment period.

CHECK	POINTS	DESCRIPTION
	3 =	More than 1 year.
	2 =	More than 6 months but less than or equal to 1 year.
	1 =	Less than or equal to 6 months.
	0 =	Not specified.

Specify follow-up period in months:

Notes:

**G. DISPLACEMENT/DIFFUSION/ANTICIPATORY BENEFITS** (if applicable) assesses the degree to which the evaluation examined for the presence of any crime displacement, diffusion of benefits, or anticipatory benefits surrounding the program implementation. (Note: This type of examination typically occurs in the evaluation of community level crime prevention efforts. The examination may involve one or many inspections and any form of displacement or diffusion, whether spatial, temporal, target, tactical, or offense.)

CHECK	POINTS	DESCRIPTION
	3 =	Central (assesses displacement as integral part of the evaluation and includes appropriate research design containing at least one treatment area, one buffer area, and one control area).
	2 =	Post-hoc (secondary assessment of displacement or diffusion with demonstration/presentation of indicators).
	1 =	Cursory (brief mention of displacement or diffusion but no demonstrated examination).
	0 =	None (displacement or diffusion effects should have assessed but were not).
	N/A	Not applicable.

Notes:

**CRIMESOLUTIONS.GOV: PROGRAM EVIDENCE RATING INSTRUMENT—PART 2**

<b>DESIGN QUALITY SCORING TABLE</b>	
	<b>Research Design Points</b>
+	<b>Sample Size Points</b>
+	<b>Statistical Adjustment Points</b>
+	<b>Instrumentation Points</b>
+	<b>Internal Validity Points</b>
+	<b>Follow-Up Period Points</b>
+	<b>Displacement/Diffusion/Anticipatory Benefits Points (if applicable)</b>
=	<b>TOTAL</b>
/	<b>NUMBER OF ITEMS (FILL IN)</b>
=	<b>DESIGN QUALITY SCORE</b>

**SCORING DIRECTIONS.** Points are summed and divided by the number of items in the dimension. *(Note: Due to the diversity in research design across program areas, some items are not appropriate for all designs. Consequently, the number of items varies by design.)*

**OUTCOME EVIDENCE**

**A. SUBSTANTIVE PROGRAM EFFECTS** rates the level of confidence that an effect is the result of the program rather than other factors (such as the selection process or by chance). **State the intent/core purpose of this program in the box on the next page.** The core purpose and primary outcomes should relate to one of the major areas of Crime Solutions (reducing crime/delinquency, improving the justice system, responding to victims, etc.). **Select and score primary and secondary outcomes (up to five each).** Secondary outcomes should relate to the ancillary purposes of the program. Scores for primary outcomes are given three times the weight of secondary outcomes. Use the following scale to assess the program’s achievement of each of the outcomes. **Be sure to focus on the core purpose and primary outcomes of the program as these are most relevant to Crime Solutions.** See example below.

POINTS	DESCRIPTION
3 =	The finding provides very strong evidence of a program effect (significant finding; large effect).
2 =	The finding provides moderate evidence of a program effect (significant finding, moderate effect).
1 =	The finding provides marginal evidence of a program effect (significant finding, small effect).
0 =	The finding provides no evidence of a program effect (comparison groups do not differ, no effect).

**EXAMPLE**

**PROGRAM INTENT/CORE PURPOSE:** The main intents of the Hawaii Opportunity Probation with Enforcement (HOPE) program are to reduce drug use, recidivism, and incarceration among probationers who are considered at high-risk of failing probation or returning to prison.

	PRIMARY OUTCOMES	FINDINGS	UNWEIGHTED SCORE	WEIGHT VALUE	WEIGHTED SCORE
Primary Outcome 1	Reduce positive urine	Comp. group had significantly higher (46%) positive urinalyses versus 13% for HOPE.	3	x 3	9
Primary Outcome 2	Reduce re-arrest rates	47% of comparison group were arrested compared with 21% of HOPE participants.	3	x 3	9
Primary Outcome 3	Reduce # days incarcerated	Participants spent an average of 48% fewer days incarcerated (138 days vs. 267 days).	2	x 3	6
	Sum			9	24

	SECONDARY OUTCOMES	FINDINGS	UNWEIGHTED SCORE	WEIGHT VALUE	WEIGHTED SCORE
Secondary Outcome 1	Reduce probation revocations	HOPE participants had 7% revocation rate compared with 15% for comparison group (statistically significant).	1	x 1	1
Secondary Outcome 2	Reduce no-shows for prob. appointments	HOPE participants were significantly (61%) less likely to skip or miss appointments than the comparison group (9% vs. 23%).	3	x 1	3
	Sum			2	4

CALCULATION WORKSHEET					
	SUM OF WEIGHTED SCORE	SUM OF WEIGHT VALUES			
Primary Outcomes	24	9			
Secondary Outcomes	4	2	<b>SUBSTANTIVE PROGRAM EFFECTS SCORE</b>		
<b>TOTAL</b>	<b>28</b>	<b>11</b>	<b>=</b>	<b>2.5</b>	

**CRIMESOLUTIONS.GOV: PROGRAM EVIDENCE RATING INSTRUMENT—PART 2**

**INTENT/CORE PURPOSE OF THE PROGRAM** *(In one sentence, state the intent/core purpose of the program.)*

**PRIMARY OUTCOMES CHART**

	<b>PRIMARY OUTCOMES</b>	<b>FINDINGS</b>	<b>UNWEIGHTED SCORE</b>	<b>WEIGHT VALUE</b>	<b>WEIGHTED SCORE</b>
Primary Outcome 1				x 3	
Primary Outcome 2				x 3	
Primary Outcome 3				x 3	
Primary Outcome 4				x 3	
Primary Outcome 5				x 3	
	<b>SUM</b>		*		

**SECONDARY OUTCOMES CHART**

	<b>SECONDARY OUTCOMES</b>	<b>FINDINGS</b>	<b>UNWEIGHTED SCORE</b>	<b>WEIGHT VALUE</b>	<b>WEIGHTED SCORE</b>
Secondary Outcome 1				x 1	
Secondary Outcome 2				x 1	
Secondary Outcome 3				x 1	
Secondary Outcome 4				x 1	
Secondary Outcome 5				x 1	
	<b>SUM</b>				

**CALCULATION WORKSHEET**

	<b>SUM OF WEIGHTED SCORE</b>	<b>SUM OF WEIGHT VALUES</b>	
Primary Outcomes			
Secondary Outcomes			<b>SUBSTANTIVE PROGRAM EFFECTS SCORE</b>
<b>TOTAL</b>		÷	=

\*If there are no secondary outcomes, the score is the average of the primary outcomes' unweighted score.



**CRIMESOLUTIONS.GOV: PROGRAM EVIDENCE RATING INSTRUMENT—PART 2**

**B. BEHAVIOR** assesses the degree to which a program demonstrates change(s) in behavior. Programs that demonstrate behavioral change (reductions in criminal behavior, substance abuse, etc.) are considered more effective than programs that demonstrate changes only in knowledge or attitudes, because behavior does not always conform to a person's feelings and beliefs. Behavior that reflects a given attitude may be suppressed because of a competing attitude, or in deference to the views of others. *(Note 1: Behavior change need not be limited to individual behavior, but may also include organizational change or changes in community-level behavior, such as an increase in convictions, a reduction in the fear of crime, or a drop in crime rates. A drop in arrests in a particular group or community may also be considered behavioral change. Note 2: Behavior change could include substantive program effects mentioned in A above.)*

CHECK	POINTS	DESCRIPTION
	3 =	The preponderance of the findings provides strong evidence of behavioral or systemic change (consistent, mostly significant findings; large effects).
	2 =	The preponderance of the findings provides moderate evidence of behavioral change or systemic (inconsistent but some significant findings, small to moderate effects).
	1 =	The preponderance of the findings provides evidence of attitudinal/knowledge change but only marginal evidence of behavioral or systemic changes (significant attitudinal findings with varying effects, but small behavioral effects).
	0 =	The findings provide no evidence of behavioral, systemic or attitudinal/knowledge change (comparison groups do not differ, no attitudinal or behavioral effect).

Notes:

---



---



---



---

**C. OUTCOME** (*directional indicator*) indicates the direction of the effects based on the preponderance of the evidence. *(Note: This element is a multiplier.)*

CHECK	POINTS	DESCRIPTION
	1 =	The preponderance of evidence indicates positive effects.
	0 =	The preponderance of evidence indicates no effect.
	-1 =	The preponderance of evidence indicates negative effects.

Notes:

---



---



---



---



---



---



**PROGRAM FIDELITY**

**A. DOCUMENTATION** refers to the process of recording information about program fidelity (i.e., the degree to which the core program services or components are implemented as designed via the program description). To effectively establish causality, program designers should operationally define the core components of the program that are necessary and sufficient to achieve the outcomes desired. The implementation of these core components should then be empirically assessed and recorded to determine if the program under study meets a minimum threshold of implementation. Program evaluation studies should then include these measures of implementation fidelity to identify the underlying casual mechanism of the program.

CHECK	POINTS	DESCRIPTION
	3 =	The collection of program implementation evidence is systematic and measured quantitatively (dosage, time spent in training, adherence to guidelines or a manual, etc.).
	2 =	The collection of program implementation evidence is systematic and assessed qualitatively (non-numeric data obtained through direct means, such as site observations, staff interviews, focus groups, etc.).
	1 =	The collection of program implementation evidence is non-systematic (ad hoc), incomplete, and/or assessed anecdotally.
	0 =	No information about of program implementation.

Notes:

---



---



---



---

**B. ADHERENCE** (*directional indicator*) refers to the degree to which the core program services or components are implemented/delivered as designed (via the program description). Adequate adherence to program design is as important as the type of program. An effective program model can be rendered less effective if implemented poorly, without fidelity. (*Note: This element is a multiplier.*)

CHECK	POINTS	DESCRIPTION
	1 =	Adherence to program appears satisfactory.
	0 =	No information about program implementation.
	-1 =	Adherence to program appears poor.

Notes:

---



---



---



---

PROGRAM FIDELITY SCORING TABLE		
	Documentation Points	
=	TOTAL	
/	NUMBER OF ITEMS	1
=	SUB TOTAL	
x	ADHERENCE: DIRECTIONAL INDICATOR	
=	FIDELITY EVIDENCE SCORE	

**Scoring Directions:** Points are summed, divided by the number of items in the dimension, and then multiplied by the directional indicator. A positive value indicates sufficient program fidelity while a negative value indicates poor program fidelity. A zero indicates that no information was provided regarding fidelity.

**REVIEWER CONFIDENCE/OVERRIDE OPTION**

The Reviewer Confidence/Override Option is intended to be used sparingly and only if the reviewer lacks confidence in the results of this scoring instrument as it pertains to the study. The Override provides an opportunity to exercise judgment and discretion based on the reviewer’s expertise for items that may not have been explicitly captured in the elements of the instrument. If the reviewer feels that no confidence can be placed in the results, detailed reasons must be provided. If this option is invoked by both reviewers, the study will be coded as a Class 5 (Insufficient Information) and will be eliminated from the review process. If one reviewer invokes the Override Option and the other does not, the dispute resolution process will be used to classify the study.

Examples of these further considerations include:

Outcomes: Study outcomes should match the intent of the program and be valid measures relating to the program’s purpose. The reviewer should take into account if the specified outcomes match the intent of the program.

Anomalous Findings: Anomalous findings may contradict the intent of the program and suggest the possibility of confounding causal variables. The reviewer should judge if anomalous findings draw into question the confidence in the results of the evaluation.

Statistical Analysis: The type of statistical analysis utilized can sometimes influence the outcomes. The reviewer should take into account whether the statistical analysis was appropriate given the research design.

Other: The reviewer should consider whether the study possesses any other limitations not expressly or inadequately addressed in the instrument that reduces the confidence in the results of the evaluation.

CHECK	POINTS	DESCRIPTION
	1 =	Confidence should be placed on the results of this evaluation because the number and type of limitations are minimal.
	0 =	Very limited or no confidence should be placed in the results of this evaluation because the number and type of limitations are too serious.*

\*Note: If “0” is selected, the reviewer must explain below why you do not have confidence in the results and why this was not captured in the scoring instrument.

---



---



---



---



---



---



---



---

**OVERALL SCORE**

	CONCEPTUAL FRAMEWORK	DESIGN QUALITY	OUTCOME EVIDENCE	PROGRAM FIDELITY
<b>Overall Score*</b>				
<p>*Reviewer Confidence/Override Option: As a final step on the scoring instrument, Study Reviewers provide an assessment as to their overall confidence in the study design. If both Study Reviewers agree that there is a fundamental flaw in the study design (not captured in the Design Quality dimension) that raises serious concerns about the study's results, the study is removed from the evidence base and not factored into the program's evidence rating. If one reviewer invokes the Override Option and the other does not, the dispute resolution process will be used to classify the study.</p>				

**STUDY CLASSIFICATION SYSTEM**

The score in each of the four dimensions is calculated separately and used to assess each study.<sup>2</sup> The maximum overall score in each dimension is 3 points. The outcome evidence and program fidelity dimensions include directional indicators to signify the directional nature of the dimension. These dimensions are then used to classify each study into one of the following five classes:

CHECK	CLASS	DESCRIPTION
	<b>Class 1 (Strong Evidence of Positive Effect)</b>	<b>This study must have exceptional scores (at least 2.0) in all four dimensions of program effectiveness.</b> In general, this study demonstrates strong evidence in favor of the program when evaluated with a design of high quality (quasi-experimental) and implemented with sufficient fidelity.
	<b>Class 2 (Some Evidence of Positive Effect)</b>	<b>This study must have above average score scores (at least 1.5) in the design and outcome evidence dimensions.</b> In general, this study demonstrates promising (perhaps inconsistent) evidence in favor of the program when evaluated with a design of high quality (quasi-experimental). More extensive research is required.
	<b>Class 3 (Strong Evidence of Negative Effect)</b>	<b>This study must have a poor score (less than 0) in the outcome evidence dimension yet exceptional scores (at least 2.0 in design and fidelity) in other dimensions of program effectiveness.</b> In general, when implemented with sufficient fidelity and using an evaluation design of high quality (quasi-experimental), this study demonstrates negative program effects.
	<b>Class 4 (Strong Evidence of Null Effect)</b>	<b>This study must have a neutral score (from 0 to 1.4) in the outcome evidence dimension yet exceptional scores (at least 2.0 in design and fidelity) in other dimensions of program effectiveness.</b> In general, this study demonstrates no evidence in favor of the program when evaluated with a design of high quality (quasi-experimental) and implemented with sufficient fidelity.
	<b>Class 5 (Insufficient Information)</b>	<b>This study must have neutral scores (less than 1.5) in design quality dimensions.</b> In general, there is insufficient evidence to rate this study. <i>(Note: Programs with only insufficient evidence will not receive an evidence rating.)</i>

<sup>2</sup> The conceptual framework and program fidelity dimensions are effect modifiers. These modifiers will not be used to exclude a program from inclusion in CrimeSolutions.gov, but will be applied as a gauge to increase confidence regarding the underlying causal mechanism of the program.

**A PROGRAM’S EVIDENCE RATING**

An aggregation of this research base is used to rate the evidence of effectiveness of each program, as follows:

Program’s Evidence Rating	Study Classification				
	Class 1 Strong Evidence of Positive Effect	Class 2 Some Evidence of Positive Effect	Class 3 Strong Evidence of Negative Effect	Class 4 Strong Evidence of Null Effect	Class 5 Insufficient Information
<b>Effective</b> Programs have strong evidence to indicate they achieve their intended outcomes when implemented with fidelity.	Must have <b>at least 1</b> study in Class 1.	May have up to 2 studies in Class 2.	Must have <b>0</b> studies in Class 3	May have <b>up to 1</b> study in Class 4	Studies do not determine Evidence Rating
<b>Promising</b> Programs have some evidence to indicate they achieve their intended outcomes.	Must have 0 studies in Class 1	Must have <b>at least 1</b> study in Class 2	Must have <b>0</b> studies in Class 3	May have <b>up to 1</b> study in Class 4	Studies do not determine Evidence Rating
<b>No Effects</b> Programs have strong evidence indicating that they had no or harmful effects when implemented with fidelity.	Must have <b>0</b> studies in Class 1	Must have <b>0</b> studies in Class 2	Must have <b>at least 1</b> study in either Class 3 or Class 4		Studies do not determine Evidence Rating